Fra Synapse til Fabric

DYNUG VÅRFESTIVAL 27. MAI 2025

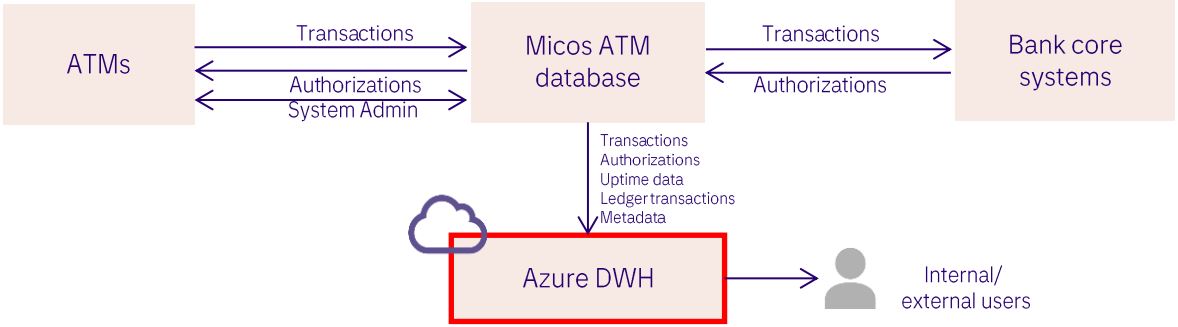
Clarion Hotel Oslo i Dronning Eufemias Gate 15

Bjørn Tingstadengen, Tietoevry

tietoevry

DWH Cards

- Project start about 2 years ago: pre-study suggested establishing a cloud based DWH for TE Banking Cards
- Tietoevry Banking Cards:
  - Cards-issuing (POS/Cards): All types transaction management and card issuing
  - Cards ATM: Outsourcing of around 11.000 ATM's, (mostly outside Norway)
  - Service center: Customer support around card services
  - Dispute: Disputed transactions/case management



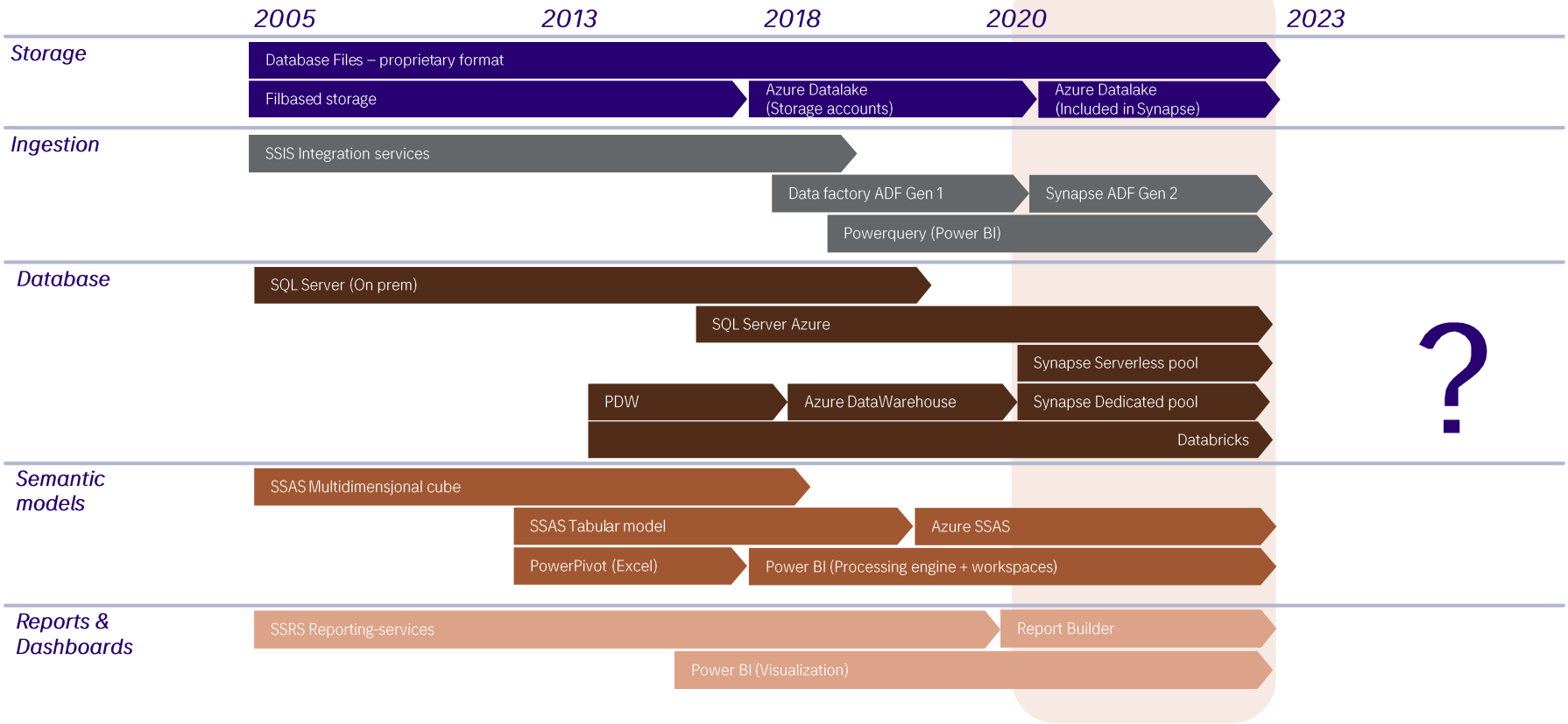
Challenges:

- Data volumes
  - ATM: 1 BN transactions per year, 1.5 BN authorizations per year
  - CAS: 3-4 BN transactions per year + x1,5 Authorizations NB! All Nordea transactions > 11-2023
  - CS & Dispute: Manual billing of customers. CD Statistics & Dispute transactions. Far less data volumes
- Standards
  - MS Azure platform preferred by Banking CIO office internally??
  - Banking ny Azure Tenants based on CAF enterprise 5.1.0 standard
  - Not fully implemented Azure plattform & principles
  - AWS Storage platform for data?

Original plan: Azure Synapse

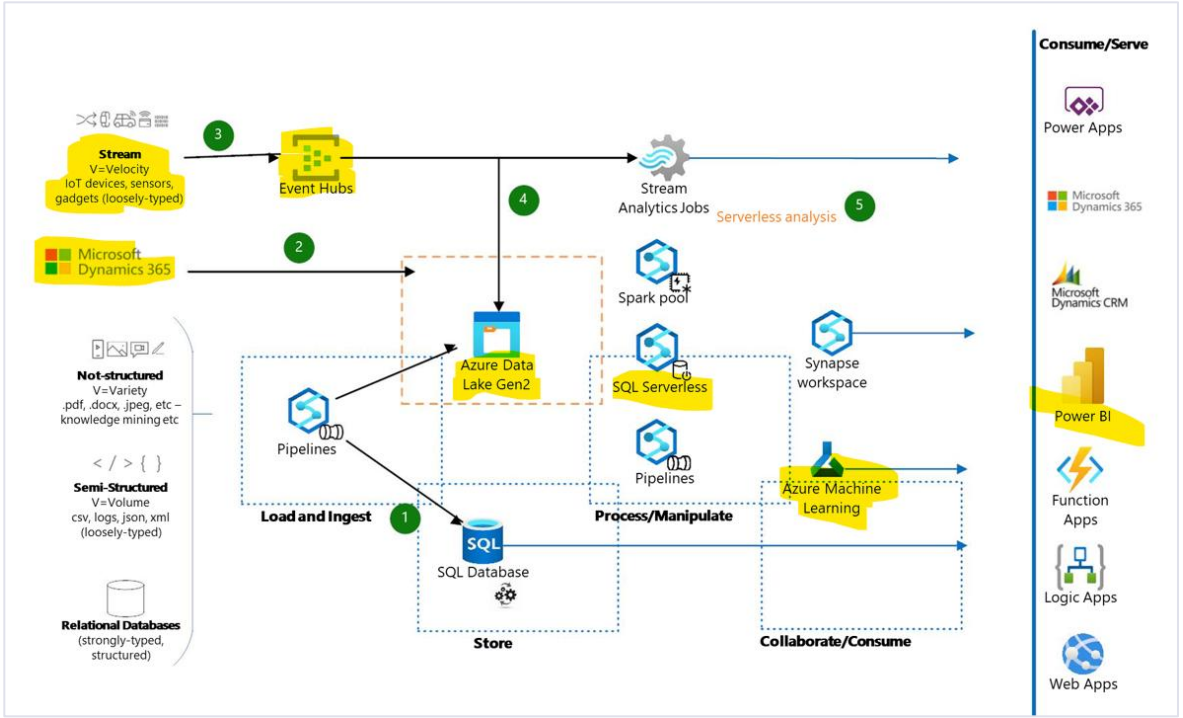
- Test load of 1 Bn rows authorization-data and 0,7Bn ATM trans-data + relevant metadata
- Testdata summarized transaction-data to Power BI import mode
- Reserve option: Snowflake

Status Microsoft BI 2020-23

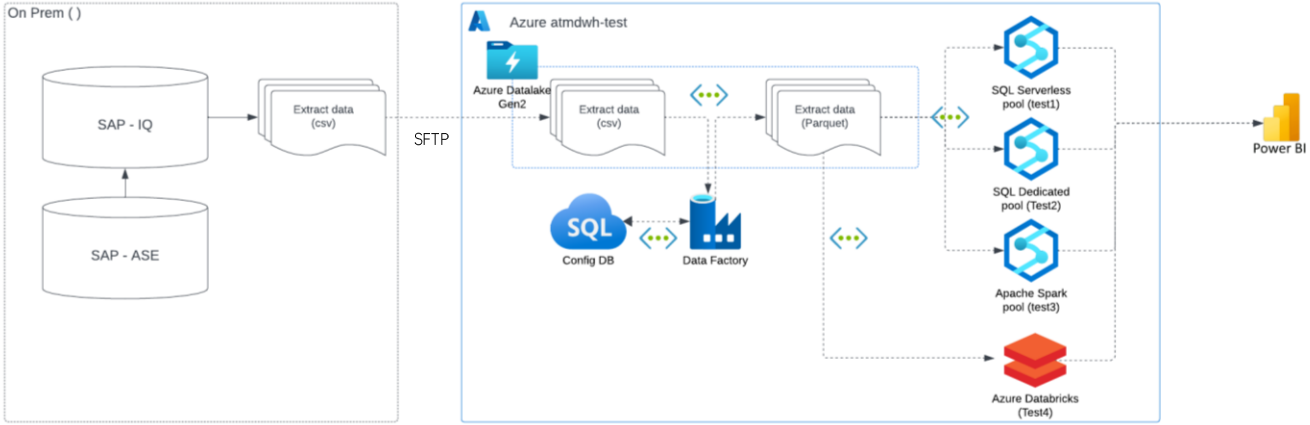


Modern Data Warehouse –Medium Business pattern

Azure SQL DB and Synapse workspace

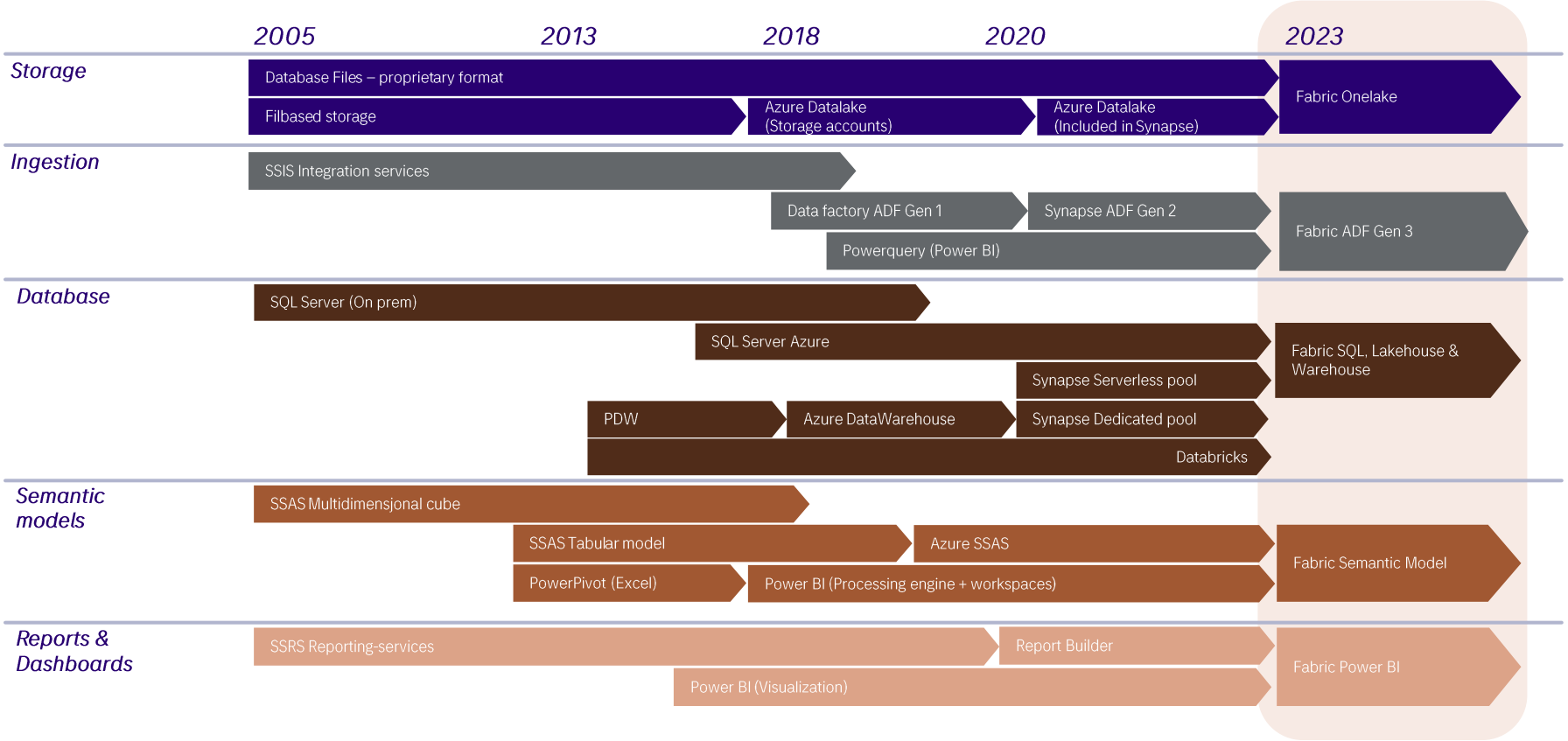


Original Azure Synapse setup

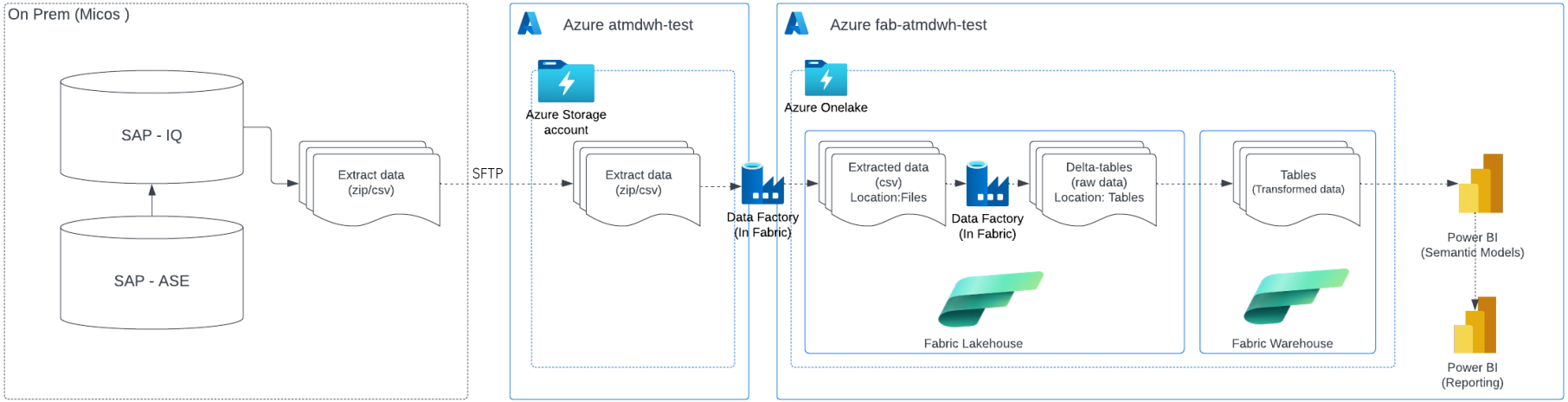


- CAF compliancy: Careful setup of V-net, components etc. to secure cloud infrastructure + Terraform (SAAS as code & infrastructure replicate) + GitHub
- SQL Serverless pool (test1): limit 30 Min execution time. Can handle up to 0,5 Bn rows if queries are not too complex! Very favorable pricing on storage only as 60 € month per TB of storage.
- Apache Spark pool (Test2): removes the 30 min limitation on the SQL Serverless pool. Solution can be scaled up. Estimated cost on a medium size installation is estimated between 0,4-0,8K €/month. Spark pool around 12 min execution time on our test-queriesRunning SQL as a Notebook needs to run ANSI SQL.
- SQL Dedicated pool (Test3): Relatively expensive on 1,1K €/month minimum. Can easily be scaled up/down to give adequate performance at a cost of around 23 €/hour. SQL Dedicated pool 5 min ex.time on testdata at (198K NOK/Month), increasing time on more complex queries. Does take advantage of T-SQL, Limited to 1 database per instance.
- Databricks (Test4): Does not take advantage of T-SQL (More ANSI like) + Python (PY-Spark). Pool Needs to be available 100 % if used as query tool. No userfriendly U/I. Advanced Security on row-level of data.

Microsoft BI 2024-



Azure Fabric



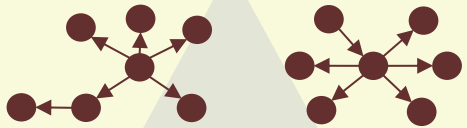
- Some Improved performance over Azure Synapse Spark pool
- Little CAF compliancy needed: pre-packed integrated components in a single resource.
- Lakehouse Serverless pool, no restrictions on size, configuration can be easily altered. NB! Power BI cost/Performance
- Starting price of around 0,4K €/mnth. Using Apache Spark pool Running SQL as a Notebook needs to run ANSI SQL not T-SQL.
- Data Warehouse New option for SQL structured data Data access data using both options SSMS/query builder (T-Sql!)
- Databricks. Not really needed?

- Could easily run incremental load of data = 10 min load time
- New Power BI SAAS option (No need for desktop client)
- Added/changed feature "Semantic models" (SSAS)
- Data Factory V3, no datasets! new dataflows (eg Power-Query), Delta-lake option
- Onelake can be included in PC filesystem (file explorer)
- No SFTP option at start of project (need separate storage account - €200/month per container)
- Some Security functionality not in place

Lagdelt arkitektur for å håndtere store datamengder, sikre datakvalitet, løsningsstruktur og tilgjengelighet

Datamarts (Gold):

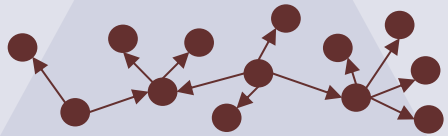
Ralph Kimball: **Bus Matrix**, Konforme Dimensioner
Bill Inmon: Star-scheme
Rapportmodeller



- **Forenklet og tilpassede data:** Dette laget inneholder data som er ytterligere bearbeidet og tilrettelagt for spesifikke forretningsbehov. Dataene er optimert for rapportering, analyse og maskinlæring.
- **Formål:** Høykvalitetsdata som kan brukes direkte i beslutningsprosesser, analyser og visualiseringer.

DataWarehouse (Silver):

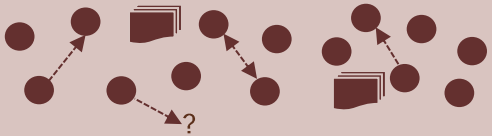
Dan Lindstedt: **Data Vault**,
Bill Inmon: **3NF**
Mike Chapple, Mike Hillyer, Fred Coulson
Relasjonsmodeler



- **Rensede data:** I dette laget blir dataene renset, transformert og normalisert. Duplikater fjernes, feil rettes, og dataene struktureres på en konsistent måte.
- **Formål:** Dette laget gjør dataene pålitelige og brukbare for videre analyse og rapportering. Kan fungere som en mellomstasjon mellom rådata og ferdigbehandlede data eller brukes direkte.

Deltalake/Lakehouse (Bronze):

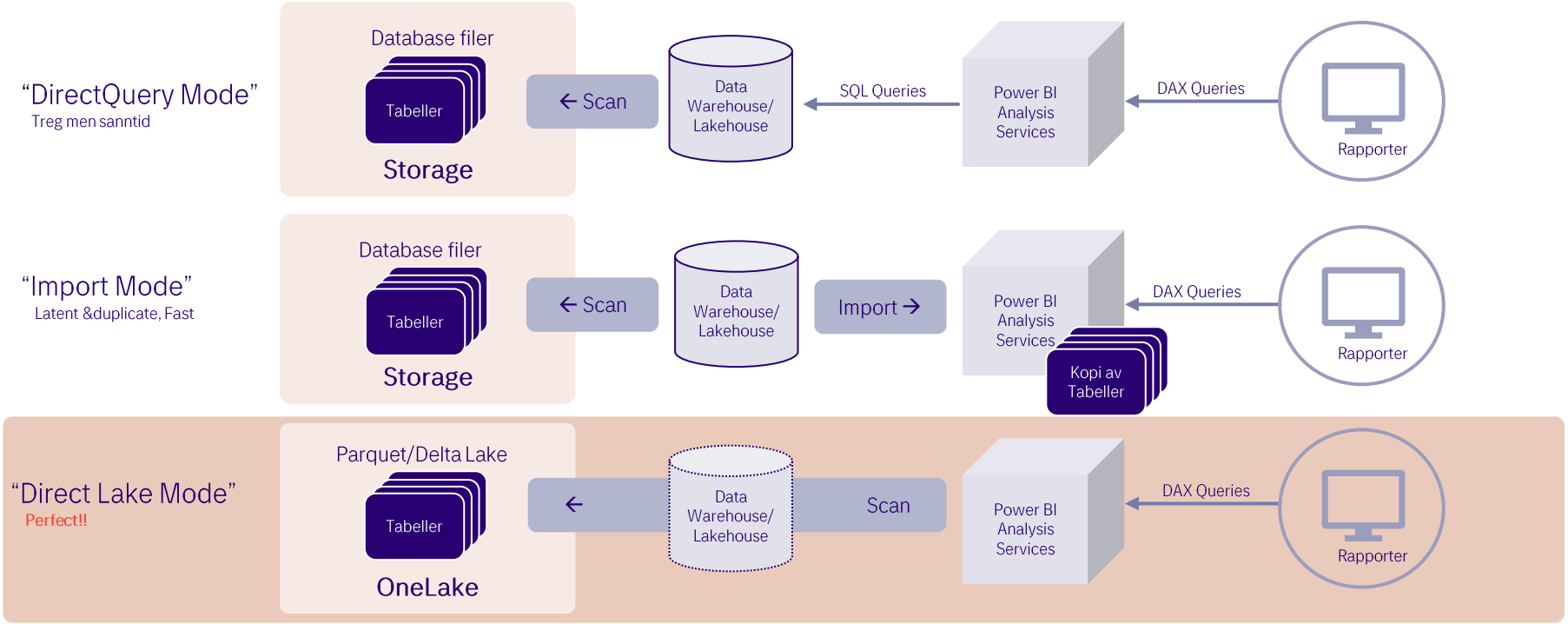
James Dixon, Liang Zhao, Bill Inmon
SQL og Ikke-Relasjonelle data
Organisert i dokumentfoldere/hierarkier,
Blobs og database-tabeller og dokumenter



- **Rådata:** Dette laget inneholder rå, ubehandlede data som er hentet direkte fra kildesystemene. Dataene er ofte i sitt opprinnelige format og kan inneholde duplikater, feil og ufullstendige oppføringer.
- **Formål:** Hovedformålet med bronse-laget er å lagre dataene slik de er, uten noen form for transformasjon eller rensing. Dette gir en historisk oversikt over alle innkommende data, sikkerhet mot feilhåndtering og mulighet til å avstemme mot kildedata.

Lagdelt arkitektur hjelper organisasjoner med å håndtere store datamengder på en strukturert måte, samtidig som de sikrer hvor databehandling gjøres. Lagring av rådata på opprinnelig format gir muligheter for å redefinere og strukturere annerledes på et senere tidspunkt om ønskelig.

Fabric –storage & query alternatives Semantic models



Azure Fabric – What we did?

The solution:

- Created DWH on ATM transactions, authorizations and ledger-data balances with necessary metadata
- Around 1,1 Bn transactions, 2 Bn bank/card authorizations and 0,9 Bn ledgerdata balances per year 0,4 Bn rows up-time data an ATM services
- Automated daily load of data
- Around 11.000 ATMs all over Western Europe (Scandinavia, Netherlands, UK, Ireland, Spain, France etc)
- 50-60 different entiteties from 7 database-instanses + some manual data as input (metadata)
- One Unified large datamodel based on Kimball's Bus-Matrix principles

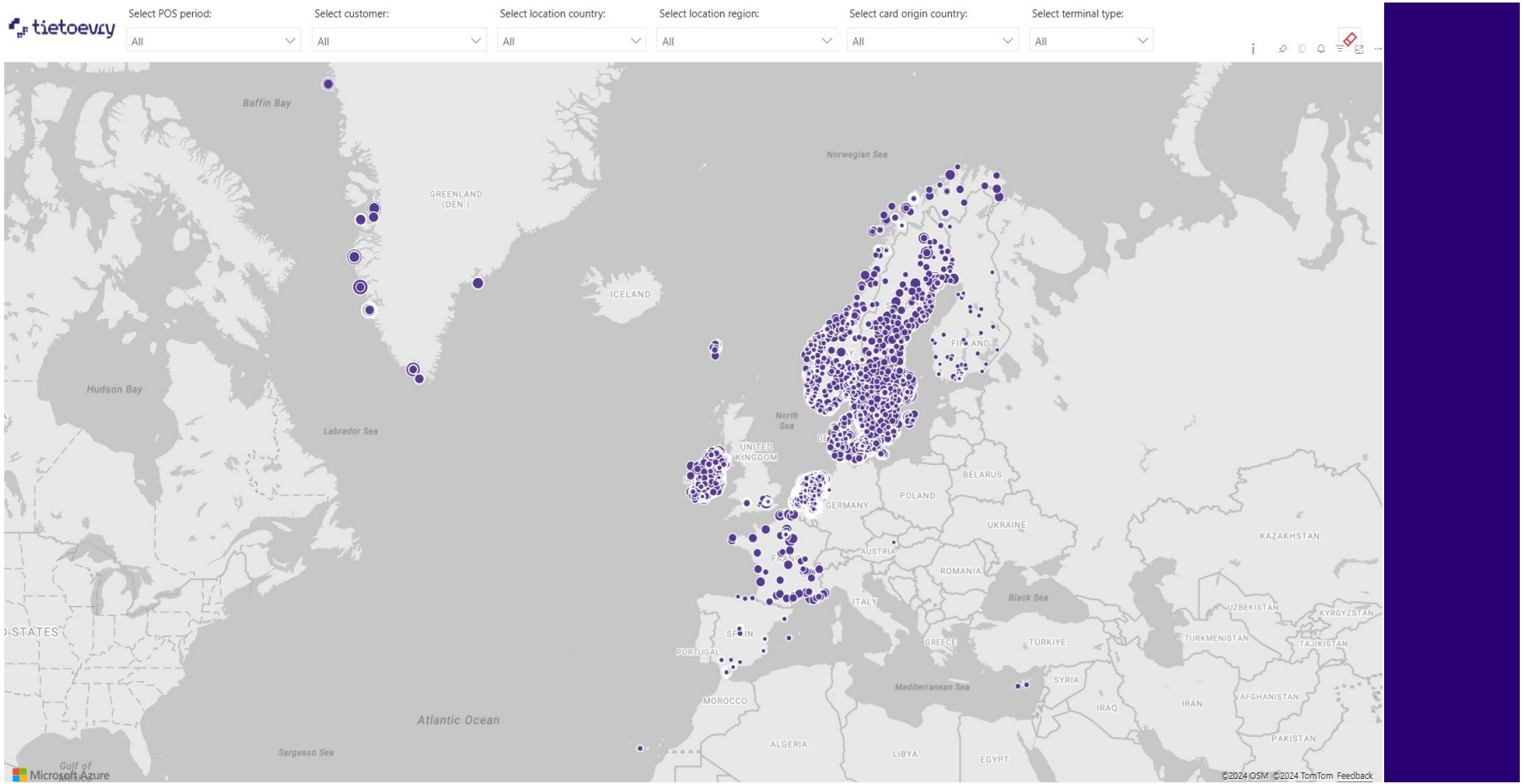
Software used:

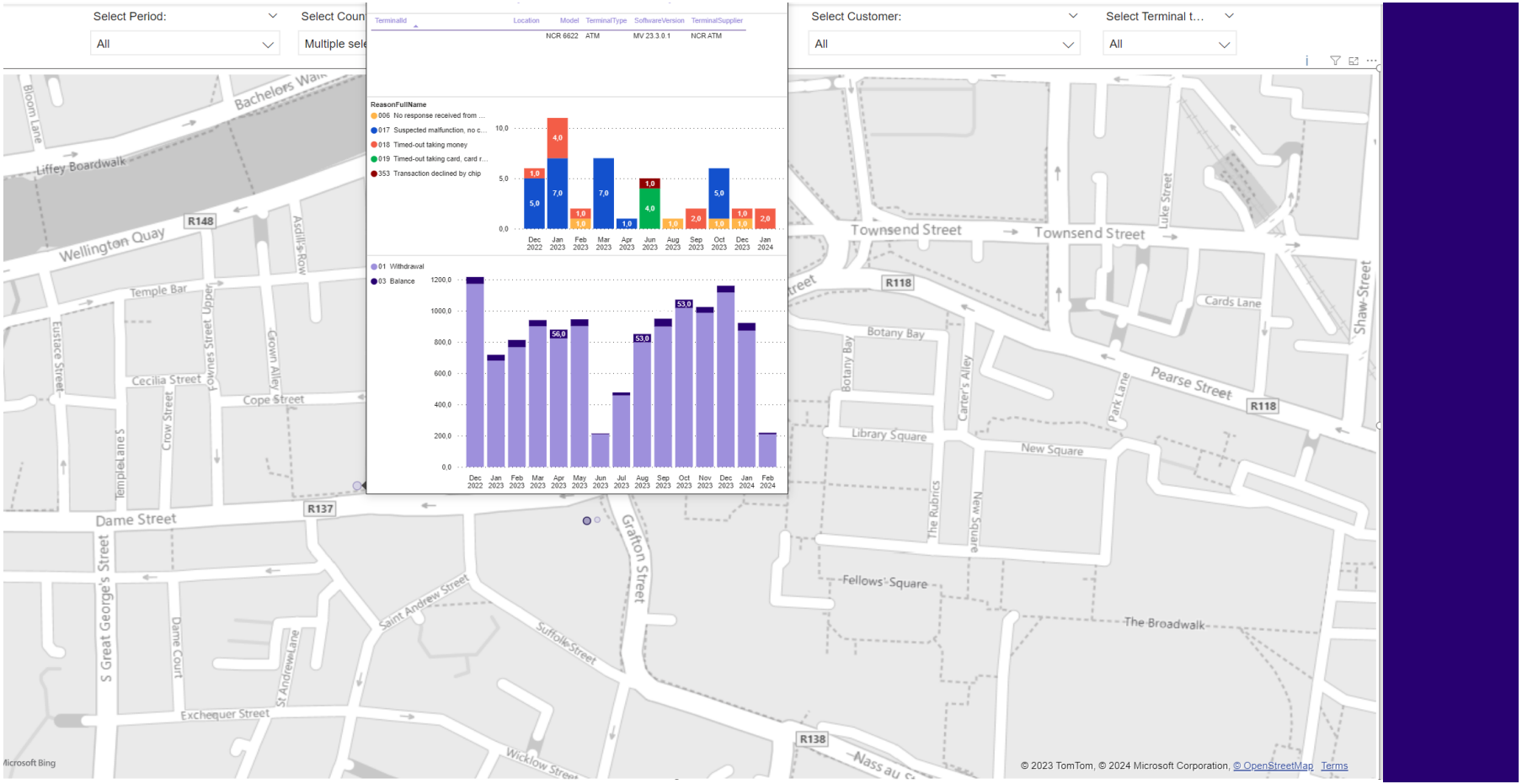
- Receiving data (all production data is stored in a secure PCI zone with very limited access): Azure Storage account
- Data movement to Fabric, convert data to parquet file-format: Azure Fabric Data factory
- Storage of filedata, both as Raw CSV data and parquet files: Azure Fabric OneLake
- Database on top of Raw non-transformed data (ODS type of data), Azure Fabric Lakehouse
- Transform data to Business-oriented values: Microsoft: T-SQL
- Database for transformed reporting data: Azure Fabric DataWarehouse database (also kind of lakehouse)
- Datamodel business reporting-layer: Fabric Semantic Model
- Reporting: Microsoft Power-BI, used in direct lakehouse query mode

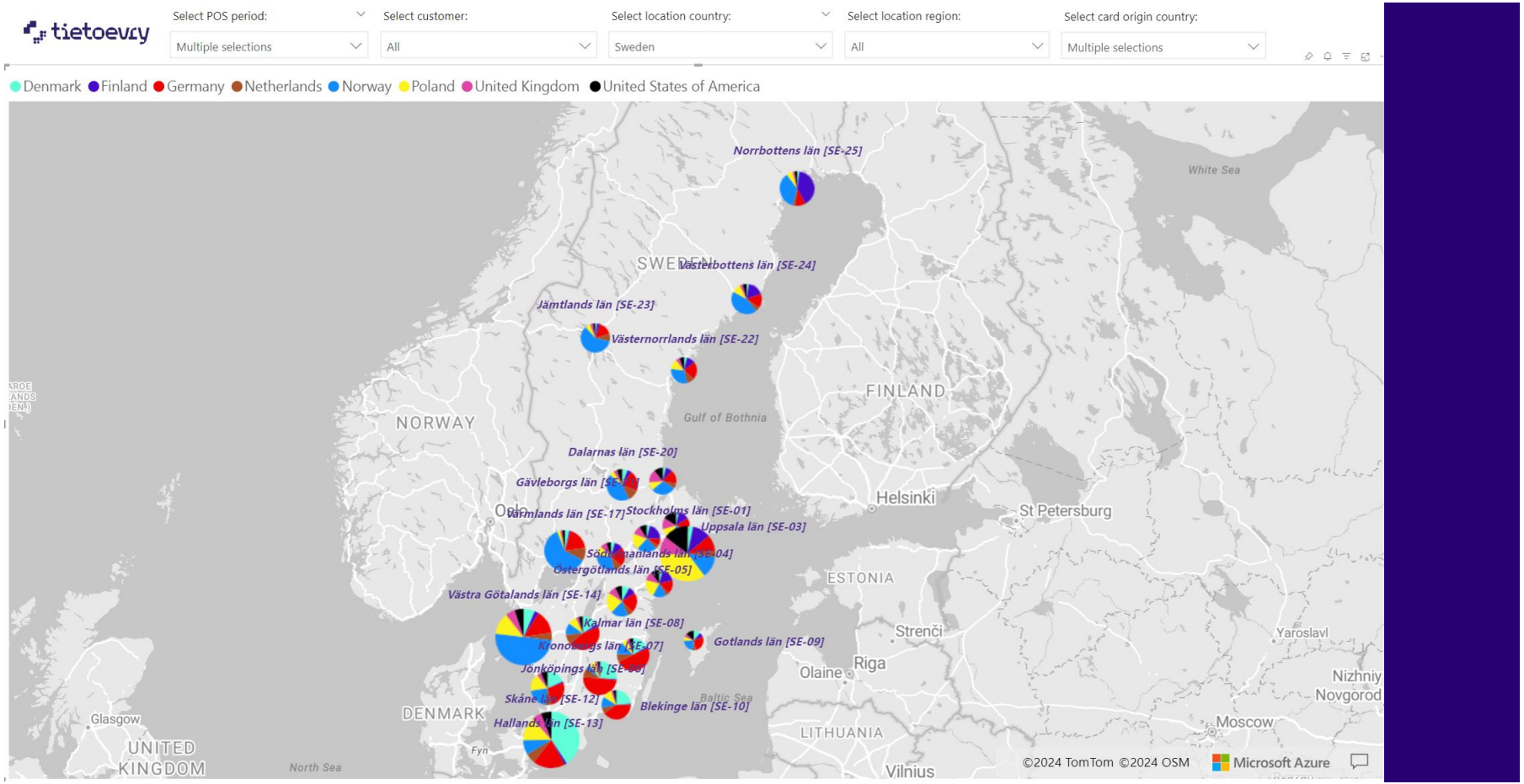
*) Parquet files is an industry-standard column-store format, which is very efficient on doing database queries on - becoming more and more commonly used

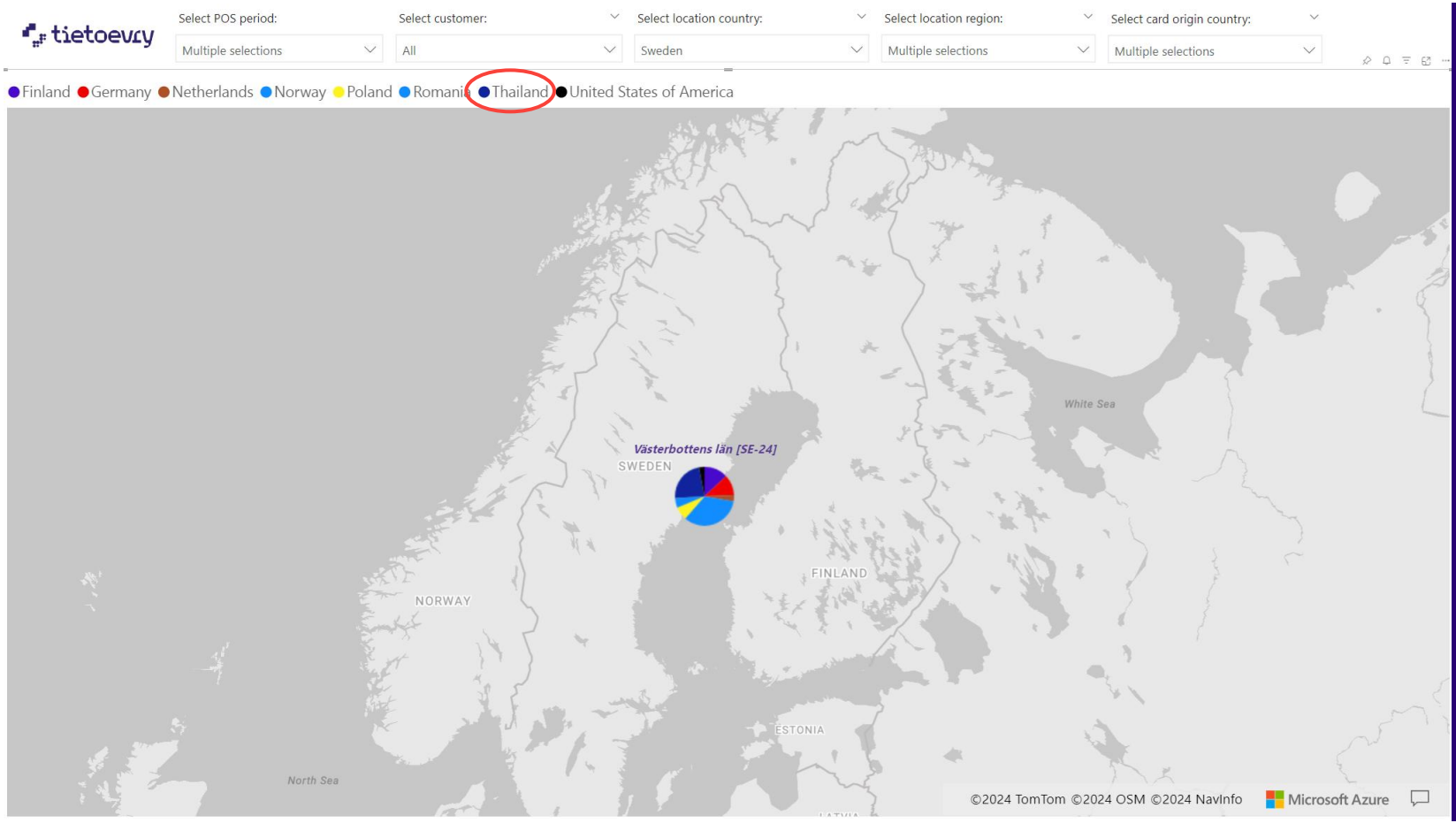
Updates new features since project start

- **Most PBI reporting works nice as SSAS** – no need for PBI desktop
- **Git integration:** More and more items can be added to Git. Since Jan 2025, introduced also folder support in Git integration, improving project organization.
- **Ownership Takeover for Fabric Items**: Enabled users to take ownership of Fabric items, facilitating better management.
- **Lakehouse Schemas (Public Preview)**: Introduced schemas in Lakehouse environments, enabling better organization and management of tables.
- **Time Travel in Fabric Warehouse**: General availability of the time travel feature, allowing users to query historical data states within Fabric warehouses.
- **Copilot:** Now available for all SKU (was only for F64 up).
- **Onelake Security is enhanced and unified**: Workspace-level security, item and folder level security, compute level security, shortcut security.
- **Mirroring**: Open for more sources, CI/CD support, firewall connectivity for Azure SQL Database
- **SQL Database (in Preview)**: fully managed, transactional database built on the same engine as Azure SQL Database.
- **New Pricing strategy: Several instances (for example prod, dev.) of Fabric in a subscription can share a common scaling SKU**
- **SFTP:** Basic FTP integration in Onelake available
- **Writeback (In Preview)**: Writeback in Power BI, on for example doing forecasts and planning (although limited process support)


tietoevry
Select POS period:
All
Select customer:
All
Select location country:
All
Select location region:
All
Select card origin country:
All
Select terminal type:
All
Baffin Bay
GREENLAND (DEN.)
Norwegian Sea
ICELAND
Hudson Bay
Labrador Sea
North Sea
UNITED KINGDOM
BELARUS
POLAND
GERMANY
UKRAINE
KAZAKHSTAN
AUSTRIA
ROMANIA
Black Sea
ITALY
UZBEKISTAN
KYRGYZSTAN
TURKMENISTAN
TAJIKISTAN
GREECE
TÜRKIYE
SYRIA
Mediterranean Sea
IRAQ
IRAN
AFGHANISTAN
PAKISTAN
MOROCCO
ALGERIA
LIBYA
EGYPT
Atlantic Ocean
Sargasso Sea
Gulf of
Microsoft Azure
©2024 OSM ©2024 TomTom Feedback


Select Period:
All
Select Coun
Multiple sele
Select Customer:
All
Select Terminal t…
All
TerminalId
Location
Model
TerminalType
SoftwareVersion
TerminalSupplier
NCR 6622
ATM
MV 23.3.0.1
NCR ATM
ReasonFullName
006 No response received from …
017 Suspected malfunction, no c…
018 Timed-out taking money
019 Timed-out taking card, card r…
353 Transaction declined by chip
01 Withdrawal
03 Balance
Liffey Boardwalk
Bachelors Walk
Bloom Lane
Wellington Quay
R148
Temple Bar
Essex Street
Temple Lane S
Crow Street
Townes Street Upper
Crown Alley
Asdill's Row
Cecilia Street
Cope Street
Dame Street
R137
S Great George's Street
Dame Court
St Andrew Lane
Saint Andrew Street
Suffolk Street
Grafton Street
Exchequer Street
Wicklow Street
R138
Townsend Street
Luke Street
R118
Cards Lane
Shaw Street
Botany Bay
Carter's Alley
Park Lane
Pearse Street
Library Square
New Square
The Rubrics
Fellows' Square
The Broadwalk
Microsoft Bing
© 2023 TomTom, © 2024 Microsoft Corporation, © OpenStreetMap Terms


tietoevry
Select POS period:
Multiple selections
Select customer:
All
Select location country:
Sweden
Select location region:
All
Select card origin country:
Multiple selections
Denmark Finland Germany Netherlands Norway Poland United Kingdom United States of America
Norrbottens län [SE-25]
Västerbottens län [SE-24]
Jämtlands län [SE-23]
Västernorrlands län [SE-22]
Dalarnas län [SE-20]
Gävleborgs län [SE-21]
Värmlands län [SE-17]
Stockholms län [SE-01]
Uppsala län [SE-03]
Södermanlands län [SE-04]
Östergötlands län [SE-05]
Västra Götalands län [SE-14]
Kalmar län [SE-08]
Kronobergs län [SE-07]
Gotlands län [SE-09]
Jönköpings län [SE-06]
Skåne län [SE-12]
Blekinge län [SE-10]
Hallands län [SE-13]
SWEDEN
NORWAY
FINLAND
DENMARK
ESTONIA
LITHUANIA
UNITED KINGDOM
Gulf of Bothnia
Baltic Sea
North Sea
White Sea
Helsinki
St Petersburg
Riga
Olaine
Strenči
Vilnius
Moscow
Yaroslavl
Nizhniy Novgorod
Glasgow
Fyn
©2024 TomTom ©2024 OSM
Microsoft Azure


tietoevry
Select POS period:
Multiple selections
Select customer:
All
Select location country:
Sweden
Select location region:
Multiple selections
Select card origin country:
Multiple selections
Finland
Germany
Netherlands
Norway
Poland
Romania
Thailand
United States of America
Västerbottens län [SE-24]
SWEDEN
NORWAY
FINLAND
ESTONIA
White Sea
North Sea
©2024 TomTom ©2024 OSM ©2024 Navinfo
Microsoft Azure

Vändningen: Bärföretag får ta hit thailändska plockare – 1200 väntas i juli

UPPDATERAD 5 JULI 2024 PUBLICERAD 5 JULI 2024

Migrationsverket avslog först alla ansökningar för thailändska bärplockare i år. Men nu har ett av besluten ändrats och Ransäters invest får klartecken från Migrationsverket.
– Vi, thailändarna, bäruppköparna, alla är glada, säger Johan Borg, vd Ransäters invest.

SVT har tidigare berättat att betydligt färre thailändska bärplockare väntas i skogarna i år. Efter missförhållanden har Migrationsverket förstärkt kontrollerna av bemanningsföretagen och nära 2000 bärplockare har fått avslag på ansökningarna under det första halvåret.

Utöver det väntar ännu 500 plockare på besked.

Lessons learnt and pricing

- Yes, it works!
- Much simplified setup
- Using direct lake mode semantic models can have a much richer set of attributes than in import mode, likewise there is less need for summarizing and simplifying data in Fabric as there is no PBI data files in the WS!
- Designing clean star-scheme models works excellent in PBI having 2+ BN rows of data
- Back-end database works smooth with 2+ Bn rows of data in a table even on SKU set to F16 (On example “select Count (*) as Antall, dim1 Group by dim1….” takes a few seconds in Azure lakehouse/Warehouse)
- With many BPI users, Front-end database needs to be scaled up – In our case to F64 (€ 9800,- /Month) + Storage cost (€23 per TB/Month)
- Dev and Prod environments share SKU cost

Desired features

- Onelake to be the receiver of SFTP files
- Connect to SFTP with SSH key (now: only allow connect with username and password)
- Improving of On-premise Data Gateway (now: work only smoothly with Dataflow)
- CI/CD: Still not optimal for deployment between workspaces in term of parameters, secrets, etc.
- Alternative selection on multiple help-screens (used as reports) in PBI visuals

Demo

Thank you

Bjørn Tingstadengen
TietoevryCreate
Snarøyveien 20, 1360 Fornebu

Bjorn.Tingstadengen@tietoevry.com
(47) 958 422 07

tietoevry